SPECIAL TOPIC

Reading Fluency as a Predictor of Reading Proficiency in Low-Performing, High-Poverty Schools

Scott K. Baker
Pacific Institutes for Research

Keith Smolkowski
Abacus Consulting and Oregon Research Institute

Rachell Katz and Hank Fien
University of Oregon

John R. Seeley
Abacus Consulting and Oregon Research Institute

Edward J. Kame'enui
National Center for Special Education Research, U.S. Department of Education

Carrie Thomas Beck
University of Oregon

Abstract. The purpose of this study was to examine oral reading fluency (ORF) in the context of a large-scale federal reading initiative conducted in low performing, high poverty schools. The objectives were to (a) investigate the relation between ORF and comprehensive reading tests, (b) examine whether slope of performance over time on ORF predicted performance on comprehensive reading tests over and above initial level of performance, and (c) test how well various models of ORF and performance on high stakes reading tests in Year 1 predicted performance on high-stakes reading tests in Year 2. Subjects were four cohorts of students in Grades 1–3, with each cohort representing approximately 2,400 students. Results support the use of ORF in the early grades to screen students for reading problems and monitor reading growth over time. The use of ORF in reading reform and implications for school psychologists are discussed.

This work was supported by an Oregon Reading First subcontract from the Oregon Department of Education to the University of Oregon (8948). The original Oregon Reading First award was granted by the U.S. Department of Education to the Oregon Department of Education (S357A0020038).

Correspondence regarding this article should be addressed to Scott K. Baker, University of Oregon, Pacific Institutes for Research, 1600 Millrace Drive, Suite 109, Eugene, OR 97403; e-mail: sbaker@uoregon.edu

Copyright 2008 by the National Association of School Psychologists, ISSN 0279-6015
Reading Fluency as a Predictor of Reading Proficiency

Over 90% of the approximately 1,600 districts and 5,283 schools in the United States that have implemented Reading First (www.readingfirstsupport.us) use oral reading fluency (ORF) to screen students for reading problems and monitor reading progress over time (Greenberg, S., Howe, K., Levi, S., & Roberts, G., personal communications, 2006). Other major education reforms, such as response to intervention (Individuals With Disabilities Education Improvement Act, 2004), have also significantly increased the use of ORF to assess reading performance. Common across these reforms is a focus on intervening early and intensively to address reading problems. An extensive research base in special education and general education provides strong support for the use of ORF as a measure of reading proficiency, but few studies have investigated the use of this measure in a nationwide federal reading initiative such as Reading First. This study addresses the use of ORF as an index of reading proficiency and as a measure of student progress over time in the context of Reading First in Oregon (http://www.ode.state.or.us/search/results/?id=96).

The roots of ORF lie in curriculum-based measurement (CBM), a set of procedures for measuring academic proficiency in basic skill areas including reading, spelling, written expression, and mathematics (Deno, 1985; Deno & Mirkin, 1977; Fuchs & Fuchs, 2002; Shinn, 1989, 1998). ORF is the most thoroughly studied of all CBM measures and has generated the most empirical support for its use. On ORF, students typically read a story or passage from grade-level reading material, and the number of words read correctly in 1 min constitutes the student’s performance score.

There is strong theoretical support for reading fluency as an important component of reading competence. LaBerge and Samuels (1974) hypothesized that automatity of reading was directly connected to reading comprehension. Based on this model of reading development, it is hypothesized that effortless word-level reading frees up attention resources that can be devoted specifically to comprehension (Adams, 1990; National Reading Panel, 2000). Posner and Snyder (1975) suggested two context-based expectancy processes that facilitate word recognition. The first consists of “automatic fast-spreading semantic activation” (Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003, p. 720) that does not require conscious attention. The second “involves slow-acting, attention-demanding, conscious use of surrounding context for word identification” (Jenkins et al., 2003, p. 720). Stanovich (1980) proposed that reading fluency results from bottom-up (print driven) and top-down processes (meaning driven) that operate concurrently when a reader confronts a word in context. Skilled readers rarely rely on conscious bottom-up processes to read words because word recognition is virtually automatic. Poor readers rely more on the context of the sentence to read words accurately because their bottom-up processes are inefficient and unreliable (Stanovich, 2000). Although there are important differences between these models, they all assert that efficient word recognition processes free up resources for comprehension. In addition, many studies have empirically demonstrated the association between ORF and overall reading proficiency, including comprehension.

ORF as an Index of Reading Proficiency

Deno, Mirkin, and Chiang (1982) published the first validity study on ORF. Five CBM measures were administered to students in special and general education (Grades 1–5). Students read words in a word list, read underlined words in passages, read words in intact passages (i.e., ORF), identified missing words in passages (i.e., cloze), and stated the meaning of underlined words in passages. ORF was the strongest measure, correlating with published criterion measures between .71 and .91. ORF correlated higher with published measures of reading comprehension than did cloze or word meaning, which were considered more direct measures of overall reading.

In a second important validity study, Fuchs, Fuchs, and Maxwell (1988) investigated CBM reading measures with middle
school students in special education. Again, a range of CBM measures were investigated included answering questions, recall, and cloze tests. ORF was the strongest measure on a number of grounds. It correlated higher with two Stanford Achievement Test subtests than the other CBM measures. It correlated higher with each of the CBM measures than any of the others. Perhaps most important, ORF correlated higher with the reading comprehension criterion measure (.92) than it did with the decoding criterion measure (.81). In other words, ORF was more strongly related to comprehension than decoding, a pattern that has been replicated in other studies (Shinn, Good, Knutson, Tilly, & Collins, 1992).

Numerous additional early studies were published establishing the validity of ORF as a measure of overall reading proficiency. One of the major conclusions of this research is that correlations between ORF and published measures of reading proficiency, including reading comprehension, are consistently moderate to strong in value, generally ranging from .60 to .90 (see Marston, 1989, and Shinn, 1998, for reviews of the research on ORF).

In the context of No Child Left Behind (2002), in which annual reading assessments are required beginning in Grade 3, a number of studies have examined the relation between ORF and performance on state reading assessments. These correlational studies have confirmed the moderate to strong association between ORF and overall measures of reading proficiency. For example, Grade 3 correlations between ORF and the reading test of the Colorado Student Assessment Program ranged from .73 to .80 (Shaw & Shaw, 2002). At Grades 4 and 5, correlations between the ORF and Colorado Student Assessment Program were .67 and .75 (Wood, 2006). McGlinchey and Hixson (2004) studied the relation between ORF and student performance on the reading test of the Michigan Educational Assessment Program from 1995 to 2002. Correlations by year ranged from .49 to .83, with an overall correlation calculated across years of .67. When ORF and the state reading tests in North Carolina and Arizona were administered in the spring of Grade 3, the correlations were .73 (Barger, 2003) and .74 (Wilson, 2005). The correlation between ORF administered in Grades 3 and 4 and the reading portion of the Ohio Proficiency Test ranged from .61 to .65 (Vander Meer, Lentz, & Stollar, 2005).

Researchers at the University of Michigan (Schilling, Carlisle, Scott, & Zeng, 2007) studied the predictive and concurrent validity of ORF with the Iowa Test of Basic Skills in Grades 1–3 in Michigan Reading First schools. ORF correlations with the Iowa Test of Basic Skills total reading score ranged from .65 to .75, and with the Iowa Test of Basic Skills reading comprehension subtest ranged from .63 to .75. Finally, Stage and Jacobsen (2001) reported correlations of .50, .51, and .51 between fall, winter, and spring administrations of ORF and the Washington Assessment of Student Learning (WASL) in Grade 4. The authors conjectured that the use of short, written answers and extended response items on the WASL, not strictly reading measures, may have led to lower correlations than usually reported involving ORF.

The consistent link between ORF and criterion measures of reading performance has been established primarily with students in Grades 3 and higher. Consequently, these studies are quite relevant in the context of No Child Left Behind (2002), in which annual assessments are required beginning in Grade 3. In Reading First, however, reading outcome assessments are also used in Grades 1 and 2 and frequently in kindergarten. Thus, it is important to understand the link between ORF and comprehensive measures of reading before Grade 3.

A specific focus of the current study is the link between ORF and specific high-stakes statewide reading tests in Grades 1–3. In this study, we refer to comprehensive measures of reading administered in Grades 1–3 as high-stakes assessments, even though for No Child Left Behind purposes high-stakes testing begins in Grade 3. However, in Reading First in Oregon and other states, comprehensive reading assessments administered at the end of Grades 1 and 2 are also used to make decisions about continued support in Reading First and other “high-stakes” decisions. When we refer
to high-stakes reading tests, we are referring to the specific tests investigated in this study and not all high-stakes reading tests.

**Oral Reading Fluency as an Index of Reading Growth Over Time**

Most research has focused on ORF as a measure of reading performance at a single point in time. Few studies have examined ORF as a direct index of reading growth over time. In the study by Deno et al. (1982), ORF performance increased as students moved up in grades, providing cross-sectional evidence of growth over time. Hasbrouck and Tindal (1992, 2006) presented normative performance information on ORF in the fall, winter, and spring in Grades 2–5 and found that as time within year and grade level increased, student performance increased. The cross-sectional data showed that students grew fastest in Grades 2 and 3. Not surprisingly, growth rates are also related to student reading difficulty. Deno, Fuchs, Marston, and Shin (2001) found that first-grade students in general education demonstrated more than twice the growth in ORF than their peers in special education.

To investigate typical growth rates for children over time, Fuchs, Fuchs, Hamlett, Walz, and Germann (1993) conducted the first longitudinal study on ORF. Different students were assessed in Grades 1–6, but in each grade the same students were tested repeatedly over time. The number of students in each grade ranged from 14 to 25. Results showed that slope of performance decreased consistently across grades. Average increases per week were 2.10, 1.46, 1.08, 0.84, 0.49, and 0.32 across Grades 1–6, respectively. These results are consistent with the cross-sectional findings reported by Deno et al. (1982) and Hasbrouck and Tindal (1992, 2006).

Higher rates of growth in Grades 1 and 2 provides support for the importance of growth on ORF by demonstrating that students who had healthy rates of growth in Grade 1 were more likely to maintain these growth rates in Grade 2 and also were more likely to end Grade 2 at grade level than students who had low rates of growth. In line with Deno et al. (2001), Speece and Ritchey (2005) also found that risk factors predicted growth on ORF in first grade. Using growth curve analysis for students at risk for reading problems at the beginning of first grade, they had predicted ORF scores at the end of the year that were less than half the magnitude of their peers not at risk (M = 20 vs. 56.9). Performance at the end of the year was based on 20 weekly ORF assessments administered from January to May. Speece and Ritchey (2005), however, did not investigate whether ORF slope was associated with performance on a strong criterion measure of overall reading proficiency.

Stage and Jacobsen (2001) investigated the value of ORF slope across fourth grade to predict performance on the WASL state test. They found that slope was not a significant predictor of WASL. However, their analysis may not have enabled a clear view of the value of slope. Stage and Jacobsen fit a hierarchical linear model of ORF that estimated an intercept in May and a slope for the preceding year. They then predicted intercept and slope with a later, end-of-year administration of the WASL. Next, Stage and Jacobsen (a) saved ORF slopes from their hierarchical linear model; (b) computed fall, winter, and spring estimates from the slopes; and (c) entered fall, winter, spring, and slope estimates into a regression model. Because they computed the fall, winter, and spring ORF estimates from the slope, the four variables in the regression likely led to multicollinearity, if not a linear dependency, inflating standard errors and yielding tests of statistical significance that are highly problematic (Cohen, Cohen, West, & Aiken, 2003).

Fuchs, Fuchs, and Compton (2004) conducted a study that examined how well level and slope on two CBMs, Word Identification Fluency and Nonsense Word Fluency, predicted performance on criterion measures of
reading, including the Woodcock-Johnson reading subtests (Woodcock & Johnson, 1989), Word Attack and Word Identification, and the Comprehension Reading Assessment Battery. Over a 1-year period, correlations between slope on Word Identification Fluency and criterion measures ranged from .50 to .85 and on Nonsense Word Fluency, the slope correlations ranged from .27 to .58. The Fuchs et al. study is similar, conceptually, to the current study because it linked slope of performance to criterion measures of reading. There are, however, two important differences between Fuchs et al. and the present study. First, the slope measure in the current study is ORF. Second, in Fuchs et al., the slope was estimated via a two-stage model and its contribution evaluated independent of students' initial performance level. The effect of slope is difficult to interpret in a model without the intercept, especially if the intercept correlates with slope (positively or negatively), as is often the case with academic tests.

In the current study, we were interested in the contribution of slope, controlling for initial level of performance. Initial level of performance on ORF, and other screening measures, is used to identify struggling readers who may require more intensive instruction. Change in performance over time is interpreted in the context of initial level of performance. The assumption is that change on ORF represents real progress that students are making in learning to read, and the degree to which students catch up with grade-level peers is based on their initial level of performance and the growth they make over time. Previous studies have not examined the degree to which change in ORF over time is actually related to better performance on specific high-stakes reading tests. Our major focus is what contribution slope makes to predicting performance on an outcome, after controlling for initial level of performance.

**Purpose of the Study and Research Questions**

Three objectives guided this study. The first was to investigate the relation between ORF and specific high-stakes reading tests for all students in Oregon Reading First. We expected the magnitude of association to be moderate to strong, consistent with prior research. The second objective was to examine whether slope on ORF predicted performance on specific high-stakes reading tests over and above initial level of ORF performance alone. Our question was, after controlling for initial level of performance on ORF in the middle of Grade 1, or the beginning of Grade 2, does growth on ORF add significantly to the prediction of performance on specific high-stakes reading measures at the end of Grades 2 and 3? Our prediction was that slope would add significantly to prediction accuracy.

The third objective was to test how well various models that included ORF and performance on specific high-stakes reading tests in Year 1 predicted performance on specific high-stakes reading tests in Year 2. In particular, we were interested in testing how well ORF stood up in prediction models that included a comprehensive measure of reading in the model. We expected that even under prediction models that included a comprehensive measure of reading, ORF would still provide important information in the prediction, consistent with the findings of Wood (2006). Thus, we wanted to know if performance on ORF were known, would performance on specific high-stakes reading tests in Year 1 contribute additional information in predicting performance on specific high-stakes reading tests in Year 2? We also wanted to know if performance on specific high-stakes reading tests in Year 1 were known, would additional information about ORF add significantly to the prediction accuracy of specific high-stakes reading tests in Year 2? We expected ORF level of performance and high-stakes tests in Year 1 to predict performance on high-stakes tests in Year 2 about equally well. We also expected ORF slope to account for additional variance in overall reading proficiency, beyond information provided by ORF level or high-stakes reading tests in Year 1.

We address Objectives 2 and 3 separately for high-stakes reading measures in Grades 2 and 3. In the results section, we
Table 1
Descriptive Data on Oral Reading Fluency and High-Stakes Primary Reading Test by Student Cohort

<table>
<thead>
<tr>
<th>Student Cohort</th>
<th>Number</th>
<th>Oral Reading Fluency Measure</th>
<th>Primary Reading Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Beginning Mean (SD)</td>
<td>Middle Mean (SD)</td>
</tr>
<tr>
<td>Cohort 4: Y 2 G 1</td>
<td>2489</td>
<td>24.13 (27.50)</td>
<td>20.54 (25.60)</td>
</tr>
<tr>
<td>Cohort 3: Y 1 G 1</td>
<td>2484</td>
<td>37.22 (30.06)</td>
<td>63.08 (38.51)</td>
</tr>
<tr>
<td>Cohort 3: Y 2 G 2</td>
<td>2417</td>
<td>32.82 (30.34)</td>
<td>58.02 (38.60)</td>
</tr>
<tr>
<td>Cohort 2: Y 1 G 2</td>
<td>2409</td>
<td>62.46 (35.55)</td>
<td>79.62 (39.63)</td>
</tr>
<tr>
<td>Cohort 2: Y 2 G 3</td>
<td>2367</td>
<td>58.44 (35.97)</td>
<td>76.54 (40.04)</td>
</tr>
<tr>
<td>Cohort 1: Y 1 G 3</td>
<td>2329</td>
<td>24.13 (27.50)</td>
<td>63.08 (38.51)</td>
</tr>
</tbody>
</table>

Note. SAT-10 = Stanford Achievement Test—Tenth Edition; OSRA = Oregon State Reading Assessment; Y = year; G = grade. Number of participants represents the number present at the fall assessment administration. Means and standard deviations for SAT-10 are scaled scores. Oral Reading Fluency is the raw score expressed as correct words per minute and the SAT-10 and OSRA are scaled scores.

address objectives 2 and 3 for students with a high-stakes reading measure in Grade 2, and then we address Objectives 2 and 3 for students with a high-stakes reading measure in Grade 3.

Method

Participants and Setting

Students from 34 Oregon Reading First schools participated in this study. All 34 schools were funded in the first cycle of Reading First and represented 16 independent school districts, located in most regions of the state. Half of the schools were in large urban areas and the rest of the schools were approximately equally divided between midsize cities with populations between 50,000 and 100,000 (8 schools) and rural areas (9 schools). In the 2003–2004 school year, 10% of the students received special education services and 32% percent of the students were English learners. Approximately 68% of the English learners were Latino students; the remaining were Asian students, American Indians, and Hawaiian Pacific Islanders.

Schools eligible for Reading First met specific criteria for student poverty level and reading performance. During the year prior to Reading First implementation (2002–2003), 69% of students across all Reading First schools qualified for free or reduced-cost lunch rates, and 27% of third-graders did not pass minimum proficiency standards on the Oregon Statewide Reading Assessment. The overall state average for free or reduced-cost lunch in 2002–2003 was 44%, and 18% of the third-grade students did not pass the third-grade test.

Data were collected during the first 2 years of Oregon Reading First implementation. Four cohorts of students participated, with each cohort representing approximately 2,400 students (see Table 1). Data from Cohort 1 were collected in Year 1 only (2003–2004) and included only students who were in Grade 3. In Year 2 (2004–2005), these students were in Grade 4, no longer in Reading First, and consequently did not provide data for analysis. Data from Cohorts 2 and 3 were collected in Years 1 and 2. Cohort 2 was in Grade 2 in Year 1 and Grade 3 in Year 2. Cohort 3 was in Grade 1 in Year 1 and in Grade 2 in Year 2. Data from Cohort 4 were collected in Year 2 only. These students were in Grade 1 in the second year of data collection.
In Year 1, students in Cohort 4 were in kindergarten and not administered ORF measures.

In Oregon Reading First, virtually all students in kindergarten through third grade participated in four assessments per year. In the fall, winter, and spring, students were administered Dynamic Indicators of Basic Early Literacy Skills (DIBELS) measures (Kaminski & Good, 1996) as part of benchmark testing. In Grades 1–3, the primary DIBELS measure was ORF. In the spring, students were administered a high-stakes reading test at the end of the year. A small percentage of students were excluded from high-stakes testing. In Grades 1 and 2, 3.3% and 3.6% of students were exempted from testing based on criteria recommended by the publisher. As with all longitudinal studies, some students failed to provide data for one or more assessments. In Grades 2 and 3, 10–13% of the students were missing data for any given ORF assessment. In Grade 1, 5% of students were missing the winter ORF assessment and 7% were missing the spring assessment. Student data were included in the analysis if they had (a) at least one ORF data point and (b) a valid score on one high stakes assessment either in Year 1 or 2. We assumed that data were missing at random and analyzed them with maximum likelihood methods that use all data available to minimize bias (Little & Rubin, 2002).

Oregon Reading First Implementation

Each Oregon Reading First school provided at least 90 min of daily, scientifically based reading instruction for all kindergarten through third-grade students with a minimum of 30 min of daily small-group, teacher-directed reading instruction. Instruction was focused on the essential elements of beginning reading (National Reading Panel, 2000): phonemic awareness, alphabetic principle, fluency, vocabulary, and comprehension. Group size, curricular programs, and instructional emphases were determined according to student instructional needs based on screening and progress-monitoring data. For example, students were carefully provided with reading material that matched their instructional level (i.e., 90% accuracy rates). Students not making adequate reading progress were provided additional instructional support beyond the 90-min reading block targeting deficient skill areas.

In each school, a Reading First mentor-coach worked closely with classroom teachers and school-based teams to support effective reading instruction. Ongoing, high-quality professional development was provided to support teachers and instructional staff. Professional development included time for teachers to analyze student performance data, plan, and refine instruction.

 Measures

DIBELS Oral Reading Fluency. The DIBELS measure of ORF was developed following procedures used in the development of other CBM measures. DIBELS ORF measures are 1-min fluency measures that take into account accuracy and speed of reading connected text. The difficulty level of the DIBELS ORF passages was calibrated for grade-level difficulty (Good & Kaminski, 2002). In the standard administration protocol, students are administered three passages at each of three benchmark assessment points during the year (beginning, middle, and end of the year) and the median score at each point is used as the representative performance score. On DIBELS ORF passages, alternate-form reliability drawn from the same level ranged from .89 to .94 and test-retest reliabilities for elementary students ranged from .92 to .97 (Good & Kaminski, 2002). In the context of Oregon specifically, the correlation between DIBELS ORF passages administered in Grade 3 and the Oregon State Reading Assessment administered in Grade 3 was calculated with 364 students and was .67 (Good, Simmons, & Kame’enui, 2001).

Test–retest reliability data have been collected on school administration of DIBELS measures, including ORF, on two occasions. In the spring of the 2004–2005 school year, six schools were randomly selected and 20% of the students in kindergarten and first grade were retested on all measures within 3 weeks of the original testing. The test–retest correla-
tion for ORF was .98, with a range of .96–.99 across the six schools. In the spring of the 2005–2006 school year, eight schools were randomly selected and approximately 20% of the students (n = 320) were retested on ORF measures in Grades 1 and 2. Mean test–retest reliabilities were .94 and .97 in Grades 1 and 2, respectively. This includes both test–retest and interrater reliability.

Stanford Achievement Test—Tenth Edition (SAT-10). The SAT-10 (Harcourt Assessment, 2002) is a group-administered, norm-referenced test of overall reading proficiency. The measure is not timed, although guidelines with flexible time recommendations are given. Reliability and validity data are strong. Kuder-Richardson reliability coefficients for total reading scores were .97 at Grade 1 and .95 at Grade 2. The correlations between the total reading score and the Otis-Lennon School Ability Test ranged from .61 to .74. The normative sample is representative of the U.S. student population.

All four of the SAT-10 subtests were administered at first grade: Word Study Skills, Word Reading, Sentence Reading, and Reading Comprehension. This entire battery takes approximately 155 min to complete. The second-grade version of the SAT-10 included the subtests Word Study Skills, Reading Vocabulary, and Reading Comprehension. The entire test takes approximately 110 min to complete.

Oregon Statewide Reading Assessment. The Oregon Statewide Reading Assessment (OSRA) is an untimed, multiple-choice test administered yearly to all students in Oregon starting in third grade. Reading passages include literary, informative, and practical selections. Seven individual subtests require students to (a) understand word meanings in the context of a selection; (b) locate information in common resources; (c) answer literal, inferential, and evaluative comprehension questions; (d) recognize common literary forms such as novels, short stories, poetry, and folk tales; and (e) analyze the use of literary elements and devices such as plot, setting, personification, and metaphor. The Oregon State Department of Education reports that the criterion validity between the OSRA and the California Achievement Test was .75 and with the Iowa Test of Basic Skills, it was .78 (Oregon State Department of Education, 2005). The four alternate forms used for the OSRA demonstrated an internal consistency reliability (Kuder-Richardson formula 20 coefficient) of .95 (Oregon State Department of Education, 2000).

Data Collection Procedures

ORF measures were administered to students by school-based assessment teams in the fall, winter, and spring. Each assessment team received a day of training on DIBELS administration and scoring. In addition, a reading coach at each school continued the assessment training by conducting calibration practice sessions with assessment team members that involved student participation. To maintain consistency across testers, the coaches conducted individual checks with each assessment team member before data collection.

The SAT-10 and the OSRA were administered in the spring. The Reading First coach supervised and monitored SAT-10 testing. Reading coaches at each school were trained by the Oregon Reading First Center. Coaches provided additional training to all teaching staff in their building on test administration and monitoring. Coaches observed testing procedures using a fidelity implementation checklist. Median fidelity on 18 test administration questions was 98.3%. Third-grade students were administered the OSRA according to procedures established by the school, district, and state. SAT-10 scoring was completed by the publisher, OSRA scoring by Oregon Department of Education. Both organizations have very strong internal structures to ensure accurate data scoring.

Data Analysis

Growth curve analyses tested how well ORF trajectories, defined by their intercepts and slopes, predicted performance on SAT-10 or OSRA administered at the end of Year 2 (Li, Duncan, McAuley, Harmer, & Smol-
Raw scores were used in all analyses of ORF data and scaled scores were used in analyses of SAT-10 and OSRA tests. For the calculation of reading trajectories over time on ORF, we used growth curve analysis to derive predicted scores in terms of change per measurement point. Descriptions of the procedures for growth curve analysis and developing the prediction models follow.

Growth curve analysis. We first used SAS PROC MIXED (SAS Institute, 2005) to construct a growth model of the repeated ORF assessments nested within individual students. The initial growth model determined the overall growth trajectories from first to third grade and included all four cohorts of students. For two of the four cohorts, measures of ORF span 2 years. For Cohort 3, there are measures of ORF from the middle of Grade 1 to the end of Grade 2, and for Cohort 2 there are measures of ORF from the beginning of Grade 2 to the end of Grade 3. We did not expect linear growth across grades because of student summer vacation. To account for this expected shift in trajectories, we added two observation-level effects that allowed a level change at the beginning of second and third grades. These terms were added to the model to improve fit and not for substantive interpretation.

We adjusted within-year growth to account for slightly greater ORF growth during the fall of second grade and a slight decline in ORF growth during the middle and end of third grade. This represents an empirically driven pattern of growth, with greater acceleration in fluency reported in earlier grades (Fuchs et al., 1993; Hasbrouck & Tindal, 1992, 2006). For the assessment at the middle of Grade 2, however, we added .2 to specify a 20% increase in growth during the first half of second grade. In the middle of third grade, we subtracted 0.2 from the linear trajectory, and we decreased the trajectory for the end of third grade by 0.4. Thus, the slope, \( T_{ij} \) (where \( i \) = each assessment occasion and \( j \) = each individual), was coded 0 in the middle of first, then 1.0, 2.0, 3.2, 4.9, 5.0, 5.8, and 6.6 by the end of Grade 3 to model the expected growth pattern.

Prediction models. We next constructed a set of models that predicted performance on comprehensive reading tests with student reading data available from 2 school years. These models compared three predictors of the performance on the comprehensive reading test administered at the end of Year 2: the ORF intercept in Year 1, the ORF slope across 2 years, and the comprehensive reading test score in Year 1. Because we used the intercept and slope as predictors, these models were fit with Mplus (Muthén & Muthén, 1998–2004), a flexible and general statistical software package built from a structural equation modeling framework.

For our prediction models, we split the sample into two groups, with one model for Cohorts 1–3, modeled across Grades 2 and 3, and another model for Cohorts 2–4, modeled across Grades 1 and 2. Figure 1 depicts the best-fitting model for Grades 2 and 3. This model shows the five observed ORF assessments that cut across Grades 1 and 2 (squares) and the ORF intercept and slope (circles). The model for grades 2 and 3 was similar in structure, except that instead of five observed ORF assessments there were six (three per grade).

The relations among the constructs of most interest are depicted in Figure 1: SAT-10 in the spring of Grade 2 predicted by (a) the ORF intercept, (b) the ORF slope, and (c) SAT-10 in the spring of Grade 1. This portion of the model has an interpretation similar to a standard regression analysis and represents the focus of this study. To evaluate the competition between predictors, we obtained standardized estimates of the regression coefficients and the variance explained in the Grade 2 SAT-10 from Mplus as the usual \( R^2 \) value. The complete model also assumes correlations between the first-grade SAT-10, ORF intercept, and ORF slope, denoted by curved lines. The Grade 2 intercept represents the level change across the summer, discussed above.

Model fit. The fit of the models to the observed data were tested with the comparative fit index (Bentler, 1990; Bollen & Long, 1993; Marsh, 1995) and the Tucker-Lewis index (Bollen, 1989; Tucker & Lewis, 1973).
Reading Fluency as a Predictor of Reading Proficiency

Criterion values of .95 were chosen for both (Hu & Bentler, 1999). We reported the $\chi^2$ value, but its sensitivity to large samples renders it too conservative as a measure of model fit (Bentler, 1990). We also provided estimates of the root mean square error of approximation (RMSEA). Values below .05 have been traditionally been recommended to indicate acceptable fit, but recent research suggests the use of more relaxed criteria (Hu & Bentler, 1999) and has criticized "rigid adherence to fixed target values" (Steiger, 2000, p. 151). Thus, we adopted .10 as our RMSEA target value for acceptable fit.

We also used Akaike's information criterion (AIC), an index of relative fit among competing models estimated with the same data (Akaike, 1974; Burnham & Anderson, 2002). The AIC was used to compare different predictor sets of the specific high-stakes tests within the same pair of grades. For the prediction of Grade 2 SAT-10, one model included the predictors ORF intercept and ORF slope, a second model included all three paths (ORF intercept, ORF slope, and Grade 1 SAT-10), a third path ORF slope and Grade 1 SAT-10, and so on. From the raw AIC value for each model, which has little meaning on its own, we computed a $\Delta$ AIC value by subtracting the AIC for the best-fitting model from the AIC for each other model. Thus, the best-fitting model necessarily has a $\Delta$ AIC of 0.0. Lower $\Delta$ AIC values indicate more support. Values of 2.0 or below indicate competitive models, and values that differ more than 10.0 from the minimum are considered to have little support over the best-fitting model (Burnham & Anderson, 2002).

Results

Table 1 presents descriptive data for ORF and the high-stakes reading tests. Average performance on the SAT-10 corresponds precisely to the 50th percentile in both Grades 1 and 2. Average performance on the OSRA corresponds to the 37th percentile and 40th percentiles in Year 1 and Year 2 of the study, respectively.

On ORF, within-year performance increased at each measurement point and across years. From the end of one grade to the next (e.g., end Grade 1 to beginning Grade 2), there is a consistent drop in performance. We attributed this drop to a summer effect and the use of more difficult reading material as students move up in grade. Finally, mean performance in relation to targeted benchmark levels of performance are typically slightly above or slightly below recommendations (Good et al., 2001). In the spring of Grades 1-3, the recommended benchmarks are 40, 90, and 110 words read correctly per min (Good et al.).

Correlations Between ORF and High-Stakes Reading Measures

Thirteen correlations between ORF and high-stakes reading tests addressed our first
research objective. Grade 1 ORF correlated .72 in the winter and .82 in the spring with the Grade 1 SAT-10. For the Grade 2 SAT-10, correlations with the five ORF assessments from winter of Grade 1 through spring of Grade 2 were .63, .72, .72, .79, and .80. Six ORF assessments from fall of Grades 2 through spring of Grade 3 correlated with the OSRA at .58, .63, .63, .65, .68, and .67. These correlations were consistent with previous research on the association between ORF and criterion measures of reading performance (Marston, 1989; Shinn, 1998).

Growth on ORF to Predict Performance on the Primary Reading Measure

To address our second objective, determining how well growth over time on ORF added to predictions of performance on the comprehensive reading measures administered at the end of Grades 2 (SAT-10) and 3 (OSRA), we began by fitting an accelerated longitudinal growth model (Duncan, Duncan, Strycker, Li, & Alpert, 1999; Miyazaki & Raudenbush, 2000) for ORF across Grades 1–3, with 11,829 students representing 38,164 ORF assessments. We tested the relative fit of several models with the AIC and chose the best-fitting model for further analyses. The best-fitting growth model included parameters for time and level adjustments for Grades 2 and 3. These effects were allowed to vary for individual students and to correlate with each other. We compared the residual variance estimate from this model to that from an unconditional baseline model of the ORF assessments with no predictors to provide an estimate of the reduction in variation in ORF assessments accounted for by the growth model (Singer & Willett, 2003; Snijders & Bosker, 1999). The growth model reduced the ORF residual variance from 1813.4 in the baseline model to 86.9 in the full growth model. Thus, the small set of fixed and random effects in the growth model accounted for 95.2% of the variance in ORF measures across time.

In predicting performance on the specific high-stakes reading test, we conducted two analyses, one with ORF data from first and second grade (Cohorts 2–4), and the second with ORF data from second and third grade (Cohorts 1–3). For each analysis, we fit six competing models and estimated their relative fit to the data. Each model used the same growth pattern of ORF described in the accelerated longitudinal analysis, but a different set of predictors of the specific high-stakes reading test were used. For second grade, we predicted performance on the SAT-10 total scaled score administered at the end of Grade 2 with the following six models: (a) ORF intercept, (b) ORF intercept and slope, (c) ORF intercept and slope and SAT-10 total scaled from Grade 1, (d) the SAT-10 administered at the end of Grade 1, (e) ORF slope and the SAT-10 from Grade 1, and (f) ORF intercept and the SAT-10 from Grade 1. For third grade, the six models predicted the OSRA at the end of Grade 3 and entailed a similar set of predictors.

The absolute fit indices, comparative fit index and Tucker-Lewis index, were greater than .95 for every model. The $\chi^2$ values were all statistically significant, which is to be expected given the large samples. All RMSEA values were below .10 and were adequate for these prediction models.

First and Second Grade (Cohorts 2–4)

ORF intercept and slope predicted a statistically significant portion of performance on the Grade 2 SAT-10 ($p < .0001$). Together, ORF level and ORF slope explained 70% of the variance on the SAT-10 high-stakes reading test at the end of Grade 2. In addressing Research Question 2, ORF slope accounted for an additional 10% of the variance on the Grade 2 SAT-10, after controlling for initial level of performance. This represents a robust contribution of slope in accounting for unique variance in the comprehensive reading measure.

Table 2 and Figure 1 give the results of research question three. The growth model using all three predictors—ORF intercept and slope across Grades 1 and 2 and the Grade 1
SAT-10—to predict performance on the SAT-10 at the end of Grade 2 fit the data best (Δ AIC = 0.0). Together, these three predictors explained 76% of the variance in SAT-10 performance at the end of Grade 2. The standardized estimates show that the first-grade SAT-10 predicts best, but ORF slope is also a strong predictor. The ORF intercept makes a statistically significant but smaller contribution, partly because it was highly correlated with performance on the SAT-10 in first grade (see Table 2).

Second and Third Grade (Cohorts 1–3)

In second and third grade, ORF intercept and slope also predicted a significant portion of performance on the third-grade OSRA (p < .0001). Together, ORF intercept and slope accounted for 52% of the variance on the OSRA. In addressing Research Question 2, slope on ORF contributed an additional 3% to prediction accuracy, which although statistically significant, represents a small unique contribution for slope.

Regarding the third research question in second and third grade, the best-fitting model also included all three predictors—ORF intercept and slope and the second-grade SAT-10 (Δ AIC = 0.0). No other models fit the data well. Together, these three predictors accounted for 59% of the variance in the OSRA at the end of third grade. The standardized path weights, shown in Table 3, indicate that most of the variance was predicted by the SAT-10 in Grade 2, and the ORF intercept predicted more variance than ORF slope. The reduced influence of slope is not unexpected because of the high correlations between

---

**Table 2**

| Grade 2 SAT-10 Score Predicted by Grade 1 SAT-10 Score and ORF Intercept and Slope Across Grades 1 and 2 |

<table>
<thead>
<tr>
<th>Variable</th>
<th>Raw Estimate</th>
<th>Standard Error</th>
<th>Std Estimate</th>
<th>t Value</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd SAT-10 intercept</td>
<td>316.85</td>
<td>(10.88)</td>
<td>.25</td>
<td>29.13</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ORF intercept</td>
<td>.42</td>
<td>(.03)</td>
<td>13.14</td>
<td>13.14</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ORF slope</td>
<td>1.86</td>
<td>(.08)</td>
<td>22.69</td>
<td>22.69</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>1st SAT-10</td>
<td>.41</td>
<td>(.02)</td>
<td>17.51</td>
<td>17.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>R²</td>
<td>.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF intercept</td>
<td>19.04</td>
<td>(.33)</td>
<td>58.04</td>
<td>58.04</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ORF slope</td>
<td>21.32</td>
<td>(.13)</td>
<td>161.19</td>
<td>161.19</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ORF Gr 2 change</td>
<td>-25.58</td>
<td>(.31)</td>
<td>-83.10</td>
<td>-83.10</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>1st SAT-10</td>
<td>533.55</td>
<td>(.57)</td>
<td>943.79</td>
<td>943.79</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd SAT-10 residual</td>
<td>491.78</td>
<td>(14.18)</td>
<td>34.69</td>
<td>34.69</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ORF intercept</td>
<td>728.32</td>
<td>(13.81)</td>
<td>52.75</td>
<td>52.75</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ORF slope</td>
<td>74.54</td>
<td>(2.26)</td>
<td>32.97</td>
<td>32.97</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ORF Gr 2 change</td>
<td>80.38</td>
<td>(9.76)</td>
<td>8.24</td>
<td>8.24</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ORF residual</td>
<td>74.80</td>
<td>(1.21)</td>
<td>61.68</td>
<td>61.68</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>1st SAT-10</td>
<td>2121.54</td>
<td>(39.03)</td>
<td>54.35</td>
<td>54.35</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Goodness of fit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>χ² (df)</td>
<td>1095.2</td>
<td>(15.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFI</td>
<td>0.973</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TLI</td>
<td>0.962</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSEA (95% CI)</td>
<td>0.091</td>
<td>(.087, .096)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Std = standardized; SAT-10 = Stanford Achievement Test—Tenth Edition; ORF = oral reading fluency; Gr = grade; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval.
### Table 3
Grade 3 OSRA Score Predicted by Grade 2 SAT-10 Score and ORF Intercept and Slope Across Grades 2 and 3

<table>
<thead>
<tr>
<th></th>
<th>Raw Estimate</th>
<th>Standard Error</th>
<th>Std Estimate</th>
<th>t Value</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSRA intercept</td>
<td>125.39</td>
<td>(3.66)</td>
<td>.21</td>
<td>34.31</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ORF intercept</td>
<td>.07</td>
<td>(.01)</td>
<td>.16</td>
<td>8.31</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ORF slope</td>
<td>.25</td>
<td>(.03)</td>
<td>18.17</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>2nd SAT-10</td>
<td>.13</td>
<td>(.01)</td>
<td>18.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>OSRA</td>
<td>.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td>ORF intercept</td>
<td>30.92</td>
<td>(3.9)</td>
<td>78.91</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>ORF slope</td>
<td>21.53</td>
<td>(1.2)</td>
<td>184.93</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>ORF Gr 3 change</td>
<td>-33.65</td>
<td>(3.4)</td>
<td>-100.22</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>2nd SAT-10</td>
<td>576.39</td>
<td>(5.6)</td>
<td>1027.72</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Variances</td>
<td>OSRA residual</td>
<td>53.35</td>
<td>(1.53)</td>
<td>34.92</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>ORF intercept</td>
<td>1071.19</td>
<td>(20.40)</td>
<td>52.52</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>ORF slope</td>
<td>53.33</td>
<td>(1.79)</td>
<td>29.74</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>ORF Gr 3 change</td>
<td>190.65</td>
<td>(13.61)</td>
<td>14.01</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>ORF residual</td>
<td>88.28</td>
<td>(1.24)</td>
<td>71.16</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>2nd SAT-10</td>
<td>2002.72</td>
<td>(39.12)</td>
<td>51.20</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Goodness of fit</td>
<td>χ² (df)</td>
<td>1743.6</td>
<td>(24.00)</td>
<td></td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>CFI</td>
<td>0.965</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TLI</td>
<td>0.959</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RMSEA (95% CI)</td>
<td>0.091</td>
<td>(.088, .095)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Std = standardized; SAT-10 = Stanford Achievement Test—Tenth Edition; ORF = oral reading fluency; OSRA = Oregon State Reading Assessment; Gr = grade; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval.

SAT-10 and ORF intercept \( (r = .78) \) and slope \( (r = .50) \), and the correlation between ORF intercept and slope \( (r = .29) \).

In summary, the best-fitting model in Grades 1 and 2 and the best-fitting model in Grades 2 and 3 included the same set of predictors: ORF intercept and slope and the high-stakes reading measure in Year 1. In both models, ORF slope accounted for a statistically significant amount of the variance in predicting the high-stakes measure. In first and second grade, the contribution of slope was greater than in second and third grade.

**Discussion**

In Grades 1–3, ORF was associated with performance on the SAT-10 high-stakes test in Grade 2 and the OSRA high-stakes test in Grade 3. Correlations ranged from .58 to .82, with most correlations between .60 and .80. This supports previous research on the association of ORF with commercially available standardized tests (Marston, 1989) as well as more recent research that has examined correlations between ORF and states’ reading tests (e.g., Barger, 2003; McGlinchey & Hixson, 2004; Shaw & Shaw, 2002; Vander Meer et al., 2005; Wilson, 2005). It also extends previous research by demonstrating positive correlations between ORF and criterion measures of reading in Grades 1 and 2.

The most important finding in this study was that ORF slope added to the accuracy of predicting performance on specific high-stakes tests in Year 2, above information provided by level of performance alone. The added value of slope occurred even when predictors in-
cluded two high-quality measures that provided unique information about level of reading performance, ORF intercept and performance on a specific high-stakes reading test in Year 1. On the SAT-10 high-stakes reading test in Grade 2, slope added an additional 10% to prediction accuracy, and on the Grade 3 OSRA high-stakes test, the added contribution of slope was 3%.

Together, ORF level and ORF slope explained 70% of the variance on the Grade 2 SAT-10. When the Grade 1 SAT-10 was added to ORF level and slope data, prediction accuracy accounted for 76% of the variance on the Grade 2 SAT-10. Although this represented a statistically significant improvement in prediction accuracy, a reasonable question is whether an improvement of 6% prediction accuracy is worth the cost and time associated with the yearly administration of high-stakes tests as a way to help predict future reading achievement.

On the Grade 3 high-stakes reading measure, ORF level and slope accounted for 51% of the variance in the OSRA. The best-fitting model included the Grade 2 SAT-10 with ORF intercept and ORF slope, and accounted for 59% of the variance in the third grade OSRA scores. This finding shows that the best-fitting model also accounted for significantly less of the variance in the high-stakes reading test than the Grade 2 fully specified model (i.e., ORF level, slope, and prior SAT-10 achievement in Grade 1).

One explanation why ORF might provide a stronger index of overall reading proficiency in Grade 2 than Grade 3 is that the nature of reading development may be different in the two grades, and the ability of reading fluency to provide an overall index of reading proficiency may diminish over this period of time. Although there is some evidence that correlations between ORF and overall reading performance decrease over time (Shinn, 1998; Espin & Tindal, 1998), these changes are typically more apparent when the grade difference is larger than 1 year. Also, previous studies have reported large correlations between ORF and criterion measures of reading proficiency in Grade 3 (Marston, 1989).

The use of a different high-stakes reading measure in third grade may have contributed to the $R^2$ reduction in predicting Grade 3 performance. In Grades 1 and 2, the SAT-10 was administered as the high-stakes measure and in Grade 3 it was the OSRA. The OSRA may measure different aspects of overall reading performance than the SAT-10, or it may be less reliable, thereby attenuating the association. As reviewed in the introduction, correlations between ORF and state reading tests range from .50 (Stage & Jacobsen, 2001) to .80 (Shaw & Shaw, 2002). McGlinchey and Hixson (2004), for example, found correlations between ORF and the Michigan state reading test to be similar to the correlations we report in this study. Stage and Jacobsen (2001) found the lowest correlations between ORF and a state test, the WASL, but they suspected that the low correlations were attributable to the use of written answers and extended response items on the WASL.

Regarding the OSRA, there is some evidence that this instrument is sound psychometrically when correlations between this measure and commercially available measures are examined. For example, the correlation between the OSRA and the California Achievement Test was .75 and the correlation between the OSRA and the Iowa Test of Basic Skills was .78 (Oregon State Department of Education, 2005). The internal consistency of the OSRA seems very strong. Four alternate forms of the OSRA demonstrated an internal consistency reliability of .95 (Oregon State Department of Education, 2000). In the current study, correlations between ORF and OSRA were largely in the .60 range. An important area of further research would be to investigate the technical aspects of state reading tests because there are many different tests being used by states to determine reading proficiency as part of No Child Left Behind.

The Importance of Growth Over Time

Practical applications of ORF growth data are extensive. For example, a standard
recommendation in Reading First schools is that ORF be administered once or twice a month for students at risk of reading difficulty. Although the current study investigated growth when ORF was administered up to three times per year, rather than every other week or monthly as is commonly recommended in progress-monitoring assessments, there is no reason to believe slope estimates generated using three times per year assessments versus more regular assessments would be substantially different. In fact, if ORF is a direct measure of reading fluency at any point in time and a moderate to strong gauge of overall reading proficiency, then slope estimates using benchmark data (e.g., three measurement probes, three times per year) should be highly correlated with progress monitoring data (single measurements biweekly or monthly).

The important point is that regular monitoring of ORF in the early grades provides data to estimate slope, and this study shows that slope is related to performance on comprehensive measures of reading, controlling for initial level of performance. Although regular monitoring of student progress on ORF has long been recommended (e.g., Shinn, 1989), no studies we are aware of have examined whether growth on ORF progress-monitoring data are related to performance on high-stakes measures of reading performance. Future studies should investigate the association between ORF slope, when administration is biweekly or monthly, and overall performance on specific high-stakes reading tests.

Methodological Considerations

We believe this study provides a potentially useful methodology for determining the value of slope on CBM-type measures. There are three important considerations. The first is indexing growth in relation to criterion measures of performance. Only a handful of studies have examined slope on ORF, and most of these suggest that steeper slopes are desirable (Fuchs et al. 1993; Speece & Ritchey, 2005). On a measure like ORF, there is inherent justification for attempts to increase slope. Developing reading fluency is an important goal on its own (National Reading Panel, 2000), and the fact that reading fluency is also associated with overall reading proficiency is an added benefit. This study indicates that increasing slope on ORF is likely to lead to better performance on comprehensive measures of reading.

A second and related consideration is to examine slope in the context of initial level of performance and without confounds. Stage and Jacobsen (2001) estimated ORF slopes and then calculated three levels of performance across the school year. In their regression model to estimate the contribution of slope, the slope and level data were not independent, likely leading to severe multicollinearity. In another study, Fuchs et al. (2004) considered slope independent of initial level of performance. That is, slopes were evaluated regardless of initial starting point, which leads to interpretation difficulties if intercepts correlate with slopes, a common occurrence in reading. We examined slope in the context of initial level of performance to better interpret its value within Reading First, where children who perform low on measures of reading fluency in the fall of second grade, for example, would be candidates for frequent progress monitoring and small-group instruction. Slope goals would be based on fall performance, and would be considered a more urgent objective for students who scored low in the fall. That is, one measure of intervention effectiveness for a student low in the fall on ORF would be to attain a slope that exceeded the slopes of other students who started high in the fall. In this way, the student would begin to catch up to the overall performance level of other students. To accurately interpret slopes, it is helpful to equate or control for the starting point.

The third methodological consideration involves the value of ORF level and slope together as a prediction package. Compared to a model that included another strong predictor, performance on a specific high-stakes reading test in Year 1, we were able to show that the model with just ORF level and slope did very well. On both second- and third-grade high-stakes reading tests administered in Year 2,
however, the strongest model included ORF level and slope and performance on a specific high-stakes test in Year 1.

By demonstrating the value of a model with multiple predictors, we are not suggesting that schools should try to use this assessment approach. We believe a strong case can be made from this study that ORF level and slope can be used to estimate how well students are doing in terms of overall reading development. It is valuable to know that a model with ORF level and slope accounts for most of the variance in predicting an outcome on a specific high-stakes test, even when other information about performance on previous high-stakes tests is available. It also seems reasonable that some schools might conclude that because of the additional value added by high-stakes test, they should be part of a comprehensive assessment framework, along with ORF level and slope. If schools have the resources, it might be useful to administer a comprehensive measure of reading performance before Grade 3.

**Implications for School Psychologists**

We believe this study has implications for school psychologists. School psychologists are highly qualified to help districts and school set up assessment systems targeting student reading performance. Determining which measures to administer, selecting a combination of measures that provide complimentary, unique information, and understanding differences between level of performance and slope, are important and complex tasks. Recent federal initiatives, such as Reading First and Response to Intervention, are pushing strongly in the direction of school-wide data collection and decision making (No Child Left Behind Act, 2002; Individuals With Disabilities Education Improvement Act, 2004). Schools are expected to have the technical knowledge to use and interpret various assessment measures for different purposes. School psychologists can assist schools in analyzing screening and growth data, for example, to determine if interventions are working for individual students and for groups of students (Shinn, Shinn, Hamilton, & Clarke, 2002).

Response to intervention provides an alternative for the identification of learning disabilities (Fuchs & Fuchs, 1998; Vaughn, Linan-Thompson, & Hickman, 2003; Individuals With Disabilities Education Improvement Act, 2004). School psychologists will be able to provide substantial assistance to schools to set up systems where the accurate measurement of learning over time takes place, and to determine whether students have received appropriate instruction that would allow them to make sufficient progress in meeting key learning objectives. In this study, level and slope of performance on ORF accounted for over 95% of the variance of ORF assessments, demonstrating that reliable growth estimates can be established in the early grades.

A centerpiece of the closer integration of general and special education services (Gersten & Dimino, 2006; Fuchs & Fuchs, 2006) will likely be the way schools measure student progress and determine whether the progress a student demonstrates in response to a specific intervention is sufficient. A protracted course of poor student progress in response to well-delivered, research-based interventions may constitute a learning disability. Under these conditions, the stakes involved in monitoring student progress and in defining adequate response to intervention are significant, and school psychologists will be expected to ensure that the approaches used in this process are valid (Messick, 1989; Gersten & Dimino, 2006; Fuchs & Fuchs, 2006).

An extension of this practice provides an opportunity for school psychologists to investigate important patterns in data such as ORF. For example, if only a few students out of many display poor reading growth in the face of what is expected to be a strong reading intervention, the implications drawn might focus exclusively on making adjustments in the reading interventions for the specific students experiencing low growth. If many students display problematic growth in the context of what is thought to be strong reading interventions, it may indicate that the source of the problem lies beyond the individual students.
In this case, poor reading growth may signal the need to examine the overall system in which reading instruction is provided while at the same time probing for immediate solutions for the problematic reading growth of individual students. This type of problem solving at both the systems level and the individual student level (Batsche et al., 2005; Tilly, 2008) is central to response to intervention (Individuals With Disabilities Education Improvement Act, 2004).

**Context and Limitations of the Study**

The context in which the study was conducted is important in three ways. First, the study was conducted in real school settings in which ORF data were used to screen students, monitor progress, and adjust instruction to meet students' needs. Second, all of the Reading First schools in this study provided highly specified reading instruction. Consequently, a great deal is known about the instructional conditions in which student reading performance and growth occurred. Third, a large number schools and students participated, increasing external validity.

This study included all students in a large-scale reading reform. It did not focus on a subset of students, such as students in special education or students at risk for reading failure. Future research should investigate relations between ORF and high-stakes tests with specific student populations, and in grades other than 1–3. Participation in Reading First is based on high poverty rates and low reading achievement. These findings are likely comparable to schools not in the Reading First program, but that is currently unknown.

Another important issue is the inability to investigate the cause of the stronger prediction of ORF in Grade 2 versus Grade 3. It is impossible to test the hypotheses in which the attenuation in variance accounted for (76% in second-grade outcomes versus 59% in third grade) is an artifact of the different high-stakes measures used in Grades 2 and 3 (SAT-10 in first and second grade and OSRA in third grade) or is attributable to developmental differences of ORF trajectories.

A final issue addresses the acceleration longitudinal growth model design. A more accurate picture of fluency development may emerge by following the same cohort of students from first grade through third grade. In other words, a cohort effect may also be accounting for some of the differences in the predictive power of ORF over time.

**Conclusions**

We believe the findings of this study support the use of ORF in the context of reading initiatives such as Reading First. In particular, ORF can be part of comprehensive assessment systems that schools develop for the purpose of making a range of decisions about students' reading. Schools are expected to identify as soon as possible students who may have or may develop reading problems, and beginning in first grade ORF can provide valuable information regarding who is on track for successful reading achievement and who is struggling. Also, the growth students make on ORF over time can be used to gauge how well students are developing reading fluency skills as well as other skills that are part of overall reading proficiency.

**References**


Date Received: August 21, 2006
Date Accepted: July 9, 2007
Action Editor: Sandra Chafouleas
Scott K. Baker, PhD, is Director of Pacific Institutes for Research. His research interests are in literacy and mathematics interventions and the instructional needs of English language learners.

Keith Smolkowski, PhD, is Associate Scientist and Research Analyst at Oregon Research Institute and Research Methodologist at Abacus Research, LLC. His professional work involves research on early literacy instruction, CBM, child and adolescent social behavior, and teacher and parent behavior management practices. His methodological work has focused on the design and analysis of group-randomized trials and the statistical modeling of longitudinal, multilevel data.

Rachell Katz, PhD, is Regional Coordinator for the Oregon Reading First Center. Her research interests include implementation of school-wide literacy programs, English language learners, and early intervention.

Hank Fien received his Ph.D. from the University of Oregon in 2004. He is currently Research Associate at the Center for Teaching and Learning, where he serves as Principal Investigator of an Institute of Education Sciences grant evaluating the impact of a read-loud curriculum on student’s vocabulary acquisition and oral retell skills. His research interests include using formative assessments to guide instructional decision making and empirically validating interventions aimed at preventing or ameliorating student academic problems.

John R. Seeley, PhD, is Research Scientist at the Oregon Research Institute and Abacus Research, LLC. His areas of interest include early intervention, serious emotional-behavioral problems, screening methodology, and advanced statistical modeling of longitudinal and multilevel data.

Edward J. Kame'enui, PhD, is Knight Professor of Education and Director of the Institute for the Development of Educational Achievement and the Center on Teaching and Learning in the College of Education at the University of Oregon. His areas of interest include the prevention of reading failure, school-wide implementation of beginning reading instruction, and the design and delivery of effective teaching and assessment strategies and systems.

Carrie Thomas Beck, PhD, is Research Associate at the University of Oregon. Her research and teaching interests are in the areas of early literacy, vocabulary instruction, and instructional design.